Where /aR/ the /R/s in Standard Austrian German?

Anke Jackschina\textsuperscript{1}, Barbara Schuppler\textsuperscript{2}, Rudolf Muhr\textsuperscript{1},

\textsuperscript{1}Forschungsstelle Österreichisches Deutsch, University of Graz, Austria
\textsuperscript{2}Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

b.schuppler@tugraz.at, anke.jackschina@gmail.com, rudolf.muhr@uni-graz.at

Abstract

The present paper investigates the conditions under which different realizations of /R/ occur in standard Austrian German. The study is based on 509 word tokens containing the phone sequence /aR/ in coda position drawn from a corpus of read speech from seven male Austrian radio speakers. Acoustic measurements of the vowel /a/ revealed that F1, F2 and F3 are significant predictors for the realization of /R/ as either trill, fricative or as absent. Moreover, /a/ tends to be longer when /R/ is absent than when it is present. Our analysis of the linguistic conditions for the different realizations of /R/ showed that /R/ is least reduced in stressed syllables and in words read in isolation. Furthermore, we observe that the segmental context significantly affects the realization of /R/. Most importantly, we find significant effects of morphology: /R/ tends to be more reduced when it is part of a grammatical morpheme than when it is part of the stem of a word. These findings inform the further development of models of pronunciation variation for human and automatic speech recognition.

Index Terms: Austrian German, /R/ reduction, morphologic effects, prosodic effects

1. Introduction

Whereas for the varieties of German spoken in Germany, pronunciation variation and phonetic detail has been given noticeable attention in the fields of linguistics and automatic speech recognition \cite{1, 2, 3}, for standard Austrian German there is a lack in speech resources as well as linguistic and phonetic studies. /R/ is one of the phonemes in German with the greatest variation for its phonetic realization (e.g., alveolar trill, uvular trill, uvular approximant or fricative, vocalizations and deletion \cite{4}). Not surprisingly, also phonetic studies about /R/ are widespread for the varieties spoken in Germany \cite{5, 6, 7, 8, 9, 10}. For the spoken standard in Austria, however, only little is known about the conditions for the different realizations of /R/ \cite{11, 12, 13, 14}. The current paper presents an auditory and acoustic investigation of the phoneme /R/ in the sequence /aR/ in standard Austrian German and analyzes the linguistic conditions (i.e., prosodic and morphological properties of the words, segmental context) under which the different acoustic realizations of /R/ are likely to occur \cite{15}.

The occurrence of the mentioned different realizations of /R/ is not only conditioned by the speech style and the social and regional background of the speakers, but also by the segmental context and the syllabic structure of the words. Whereas /R/ is mostly vocalized as /\textipa{r}/ after vowels in coda positions, after /a/ it has been observed to be frequently reduced and deleted, also in carefully pronounced read speech. For the current study, we chose to analyze coda /R/ following /a/, since on the one hand /R/ is highly variable in this context, but on the other hand /a/ has been reported to show very little variation in standard Austrian German. Iivonen \cite{16}, for instance, reported that the difference between tense and lax /a/ is only given by their duration. He found that whereas for German speakers F1 and F2 differ significantly for short and long /a/, for Austrian speakers, F1 and F2 are nearly identical.

Previous studies on the realization of /R/ in standard Austrian German are very limited. Ulbrich, H. \cite{7}, for instance, studied the speech from 45 male and female German news readers and actors. He found that the fricative [r] is frequently realized after /a/, and that the absence of /R/ after /a/ is only marginal in standard German. For standard Austrian German, Klaß\cite{13} reported that the most common realization of /R/ after vowel is the vocalization. Both of these studies are based on auditory phonetic analysis only. In contrast to these studies, we base our analysis on more heterogenous data and we carry out acoustic measurements of /R/ and its preceding vowel /a/.

Another relevant study on standard Austrian German has been presented by Ulbrich \cite{12}. Based on read speech from radio speakers from Austria, Switzerland and Germany, she found that the duration of stressed syllables is higher than the duration of unstressed syllables (a finding also reported in \cite{15}).

The study of Spiekermann \cite{10} shows that the changes in formants of the nucleus of an accented syllable depend on the place of articulation of the surrounding consonants. Also in the present study, we will investigate the effect of the realization of /R/ on the formants of the preceding vowel and whether these effects are different in stressed and unstressed syllables.

Recently, several studies on Germanic languages have reported effects of the morphological properties of a word on its phonetic realization (e.g., for English \cite{17, 18}, for Dutch \cite{19, 20}, for German \cite{21}). For instance, Schuppler et al. \cite{22} found that /R/ tends to be more often absent when being part of the stem of a word than when forming a grammatical morpheme. They argue that a stem is derived from the mental lexicon as a whole and thus is more likely to be reduced than grammatical morphemes, which need to be added to a stem. In the current study, we will investigate whether the realization of /R/ differs in tokens where it is part of the stem of a word compared to tokens where it is part of a grammatical morpheme.

The aims of the present paper are: (1) to study the effect of the linguistic properties of the words on the different realizations of /R/ in the phone sequence /aR/ in standard Austrian German, and (2) to investigate how these different realizations of /R/ affect the acoustic properties of the preceding vowel /a/ (F1 - F3, duration). In the following sections, we will describe the material used and the auditory and acoustic phonetic analyses performed. Subsequently, we will present our results and discuss them in the light of previous findings from the literature.
2. Materials and method

2.1. Speech material
We extracted word tokens containing the sequence /aR/ (with /R/ in syllabic coda position) from two sources: First, we extracted 468 tokens from the ADABA corpus (“Austrian Pronunciation Database” [23, 24], which contains studio recordings of read speech produced by one male trained speaker of standard Austrian German. Second, we extracted 42 tokens from read speech produced by six male Austrian radio announcers. We extracted these tokens from a corpus, which consists of 245 sentences produced in 30 minutes of speech by six male and eight female narrators from three Austrian radio stations (i.e., ORF1, ORF Steiermark and Antenne Steiermark). As all speakers were similarly trained, the speech material forms a representative basis to study standard Austrian German. In total, we have 509 tokens from 486 different word types for our analysis.

2.2. Auditory and acoustic analysis
The first author of this paper listened to all tokens once without inspection of the spectrogram and another two times with the acoustic signal and spectrogram. Each time she annotated /R/ as either produced as a trill, a fricative, deleted or as deleted, but with a perceptual lengthening of /a/. An apico-alveolar trill was observed in one token only, which was thus excluded from the analysis. Furthermore, she annotated for each token whether it carried word stress (primary, secondary, none).

Acoustic analyses of all tokens were carried out using the software PRAAT [25]. For each token, a TextGrid with interval tiers for word and phone level were created. Boundaries were set following the guidelines provided in [26] and they were moved to the nearest zero crossing. In the case of preceding voiceless segments, the first positive zero crossing was determined as the start of the vowel [15]. Based on these boundaries, segment and word durations were calculated. Also for the measurement of the formants, we followed [15], whose method is also suitable for short vowels, which sometimes do not show any steady state portion. The average formant values for /a/ preceding /R/ in our data were: F1 = 694 Hz, F2 = 1343 Hz and F3 = 2481 Hz. These values are similar to what has previously been reported for /a/ in spoken German. Ivonen [16, 11], for instance, reported for male German speakers a mean of F1 of 700 Hz and a mean F2 of 1300 Hz. Our average F3 lies in the range of the values given in [15] (F3_{min} = 2014, F3_{max} = 2783).

2.3. Statistical analysis
In order to estimate the effects of the different predictors for the different realizations of /R/, we fitted logistic regression models with a binomial logit link function and contrast coding using the statistical software R [27]. We followed the following procedure: First, we built a control model with the different realizations of /R/ as dependent variable and with the prosodic variables as independent variables. To this model, we separately introduced predictors and interactions were removed from the models.

3. Results and discussion
On the basis of the auditory analysis of the complete data set, we observed that in 23.0% of all /aR/ tokens /R/ was absent. In detail, 58% of the tokens were produced as a trill, 19% were produced with a fricative, 8% were completely deleted and 15% were deleted, but the preceding /a/ was perceptually longer.

The following analysis was motivated by two main questions: (1) Which linguistic properties predict the different realizations of /R/ in the phone sequence /aR/ in standard Austrian German (absence versus presence of /R/; /R/ realized as fricative vs. trill)? (2) How do these different realizations of /R/ affect the preceding vowel /a/?

3.1. Presence versus absence of /R/

3.1.1. Control Model (CM1): Effects of prosody
The independent variables of the Control Model (CM, cf. Table 1) were: the prosodic variables Stress (52.1% of the tokens carried primary syllabic stress, 29.1% secondary stress and 18.8% without stress) and Number of Syllables in the word (ranging between one and eight syllables, with a mean of 2.9). Furthermore, we added the variables Position in Word, which can have the values word-initial-syllable (57.2%), word-medial-syllable (22.2%) and word-final-syllable (20.6%). As several studies have shown that function words tend to be more reduced than content words (e.g., [28]), we included the independent variable Word Class (function word or content word). Finally, we also added the variable Isolated, which indicated whether words were read in isolation or within a sentence. The CM was calculated for the complete data set (N = 509).

The results for the CM are shown in Table 1. We find that the prosodic variable Stress is a significant predictor for the absence versus presence of /R/. /R/ is significantly more often deleted after long vowels in unstressed syllables (46.9%) than in secondary (19.6%) and primary stressed syllables (14.3%). These results are in line with previous results on other Germanic languages. For instance, for Dutch it has been shown that /R/ is significantly more often deleted after long vowels in unstressed syllables (30%) than in stressed syllables (23%) [29]. Furthermore, for English it has been shown that, in general, stressed syllables tend to be longer than unstressed syllables [30].

Our results also show that /R/ is significantly more often deleted in word-final position (48.6%) than in word-medial (17.7%) and word-initial position (14.1%) and that significantly more deletions of /R/ in words that were read within a longer sentence (63.4%) than for words read in isolation (18.4%). This result is as expected, since one can assume that when words are read in isolations, all segments tend to be clearly pronounced and coarticulation and reduction is less pervasive.

3.1.2. Effects of morphology
In order to analyze the effect of the morphological properties of the word on the realization of /R/, we added the independent variable Morphology to the control model (CM1). Morphology distinguishes word tokens where /aR/ is part of a grammatical morpheme (e.g., wunder-bar ‘wonderful’) and tokens where /aR/ is part of the stem of a word (e.g., Formular ‘form’). Morphology resulted to be a significant predictor ($\beta = 1.58, z = 3.69, p < .0001$): /aR/ is significantly more often produced as /a/ when occurring in a grammatical morpheme (62.1%) than when being part of the stem of a word (18.9%). In order to test whether this effect of morphology was also true for the word-final /aR/ sequences, we built a model on only those tokens, where /aR/ occurred in word-final position (N = 105).

Also on this subset, the effect of Morphology resulted to be significant ($\beta = 2.09, z = 3.74, p < .0001$).
Table 1: Control Models: CM1 for absence vs. presence of /R/ and CM2 the realization of /R/ as fricative [ɾ] or trill [ɽ].

<table>
<thead>
<tr>
<th>Absence vs. presence of /R/</th>
<th>Realization of /R/ as fricative [ɾ] or trill [ɽ]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control model CM1</strong></td>
<td><strong>Control model CM2</strong></td>
</tr>
<tr>
<td>( N = 509 )</td>
<td>( N = 397 )</td>
</tr>
<tr>
<td>Intercept</td>
<td>Intercept</td>
</tr>
<tr>
<td>-1.71</td>
<td>-0.30</td>
</tr>
<tr>
<td>( \beta ) z-value p-value</td>
<td>0.46 0.53  &lt;.0001</td>
</tr>
<tr>
<td>Stress ‘secondary’</td>
<td>Stress ‘secondary’</td>
</tr>
<tr>
<td>-0.21</td>
<td>-0.90</td>
</tr>
<tr>
<td>0.33 1</td>
<td>0.27  &lt;.0001</td>
</tr>
<tr>
<td>Stress ‘none’</td>
<td>Stress ‘none’</td>
</tr>
<tr>
<td>-1.32</td>
<td>-0.92</td>
</tr>
<tr>
<td>0.33  &lt;.0001</td>
<td>0.34  &lt;.0001</td>
</tr>
<tr>
<td>Isolated ‘yes’</td>
<td>Isolated ‘yes’</td>
</tr>
<tr>
<td>2.56</td>
<td>1.91</td>
</tr>
<tr>
<td>0.39  &lt;.0001</td>
<td>0.56  &lt;.0001</td>
</tr>
<tr>
<td>Position Word ‘initial’</td>
<td>Position Word ‘initial’</td>
</tr>
<tr>
<td>1.63</td>
<td>1.63</td>
</tr>
<tr>
<td>0.31  &lt;.0001</td>
<td>0.31  &lt;.0001</td>
</tr>
<tr>
<td>Position Word ‘medial’</td>
<td>Position Word ‘medial’</td>
</tr>
<tr>
<td>1.26</td>
<td>1.26</td>
</tr>
<tr>
<td>0.34  &lt;.0001</td>
<td>0.34  &lt;.0001</td>
</tr>
</tbody>
</table>

These results are in contrast to the findings presented in [22]. They have found that /ɾ/ tends to be more often deleted when being part of the stem of a word than when being a grammatical morpheme. They argue that a stem is derived from the mental lexicon as a whole and thus more likely to be reduced than grammatical morphemes which are added to a stem. This reasoning is only partially true for the current study. Whereas in [22] exclusively inflectional morphemes were analyzed (where speakers need to choose between an inflectional paradigm), most of our grammatical morphemes were suffixes (i.e., -bar), which are used to create adjectives (i.e., where the choice of suffix does not depend on the uttered sentence). In addition, German adjectives can easily be distinguished from nouns by their position in the sentence. Thus, we suggest that the higher degree of reduction of /ɾ/ in grammatical morphemes in our German data can be explained by the lower information load of grammatical morphemes compared to the stems of words.

3.1.3 Effects of segmental context

In order to investigate the effect of the preceding context, we fitted a model on those tokens, where /aR/ was not word-initial (\( N = 436 \)). The independent variables added to the control model were Morphology, Preceding Manner (fricative, nasal, plosive and rest), Preceding Place with the values front (i.e., bilabial, labiodental), mid (i.e., alveolar, palatal) and back (velar, uvular), and Preceding Voiced, which indicates whether the preceding consonant is voiced. Preceding Manner resulted to be the only significant factor: /aR/ is significantly less likely to be realized as /a/ after nasals (11.9%) and fricatives (15.1%) than after plosives (29%); \( \beta = -1.44, z = -2.45, p < .01 \) and rest (46.0%); \( \beta = -2.54, z = -3.36, p < .0001 \).

For the analysis of the effect of following context, we fitted a model on those tokens, where /aR/ was not word-final (\( N = 429 \)). The independent variables added to the control model were Morphology, Following Manner, Following Place and Following Voiced, with the same values as described for the preceding context. Similar as for the preceding context, we observe effects of Following Manner. The sequence /aR/ is significantly less probable to be reduced to /a/ before plosives (6.8%, \( \beta = 3.32, z = 2.96, p < .0001 \)) than before nasals (13.3%), fricatives (29.8%) and rest (52.3%). These findings are in line with the counts of /R/s presented by [13], who mentioned that most instances of /R/ were realized before plosives.

3.1.4 Acoustic measures of the preceding vowel /a/ as predictors for /R/ deletion

As described in Section 2.2, we measured the duration and the formants F1-F3 of the vowel /a/ which precede /R/. The aim was to investigate whether /a/ differs acoustically in tokens where /R/ is absent from tokens where it is not absent. For the statistical analysis, we added the measures Duration (in ms), F1, F2 and F3 (in Hz) to a model with all significant predictors presented in the previous sections. Duration showed to be a significant predictor (\( \beta = -28.4, z = -5.33, p < .0001 \)) and to have a significant interaction with Stress (Interaction Duration and Stress: \( \beta = -38.3, z = -2.76, p < .01 \)); the duration of /a/ is significantly longer when /R/ is absent than when it is present, and this effect is stronger for stressed than for unstressed syllables. Figure 1 shows the durations of /a/ for the different realizations of /R/. Already [4] found similar tendencies for the durations of vowels preceding /R/ in German, but only for several tokens. Our study now provides evidence based on a larger database.

Our analysis also showed that F1 is significantly lower (\( \beta = 0.01, z = 4.32, p < .0001 \)) when /R/ is absent (\( F_{\text{mean}} = 657 \) Hz) than when it is realized as either fricative or trill (\( F_{\text{mean}} = 704 \) Hz). F2 is higher when /R/ is absent (\( F2_{\text{mean}} = 1,350 \) Hz) than when it is present (\( F2_{\text{mean}} = 1,341 \) Hz), and this effect of F2 is significantly strongest for syllables carrying primary word stress (Interaction Stress and F2: \( \beta = -0.01, z = -2.11, p < .001 \)). Finally, our data shows that F3 is a highly significant predictor (\( \beta = -0.01, z = -4.35, p < .0001 \)): /a/ tends to have higher F3 (\( F3_{\text{mean}} = 2,608 \) Hz) in tokens where /R/ is absent than where it is present (\( F3_{\text{mean}} = 2,445 \) Hz). F3 has been shown to be responsive to front versus back constriction. For instance, [15] mentions for pre-palatal vowels that a rise in F3 is characteristic for a palatalization of pre-palatal vowels. In our case of the pharyngeal vowel /a/ [15], the rise in F3 could be interpreted as a result of an approximation of the tongue to the place of articulation of the not-realized /R/.

![Figure 1: Duration of /a/ preceding /R/ either realized as fricative, vibrant, deleted or deleted with a perceptually longer /a/](image-url)
In general, our findings concerning the acoustic properties of /a/ preceding /R/ are in line with the results by [31]. He studied word-pairs for which the meaning of the word changes when /R/ is absent (e.g., start ‘start’ and staut ‘state’). He found that /a/ is longer in words where /R/ was absent and that the vowels are qualitatively different in spite of the identical segmental context.

3.2. Realization of /R/ as either fricative [r] or trill [n]

3.2.1. Control Model (CM2) : Effects of prosody

For our analysis of predictors for whether /R/ in the sequence /aR/ was realized as fricative [r] or trill [n], we first fitted a control model (CM2, cf. Table 1) on all word tokens where /R/ was not deleted (N = 397). Overall, 25.2% of the tokens in this data set were produced as [r] and 74.8% as [n], respectively. We included the same independent variables as described in Section 3.1.1. In contrast to our results on the complete data set, Position in the Word appears not to affect the realization of /R/ as either fricative or trill. The independent variables Isolated and Stress, however, resulted to be significant predictors: /R/ is more likely to be realized as [n] in words read in isolation (40.0%) than in words embedded in longer sentences (76.2%) and /R/ is more likely to be realized as [n] in stressed syllables (76.2%) than in unstressed syllables (66.7%).

3.2.2. Effects of morphology

As a next step, we added the independent variable Morphology to CM2, which resulted to be a significant predictor (β = 2.08, z = 3.18, p < .001): /R/ is more likely to be realized as [n] when /R/ is part of the stem of a word (76.5%) than when it is a grammatical morpheme (28.6%). These results go into the same directions as our results concerning the morphological effects on the absence vs. presence of /R/. We conclude that in grammatical morphemes reduced realizations of /R/ (i.e., [r] or deleted) are more likely than in word stems.

3.2.3. Effects of segmental context

In order to investigate the effect of the preceding context, we fitted a model on those tokens, where /aR/ was not word-initial (N = 324). Preceding Place resulted to be the only significant factor: /R/ is significantly more likely to be realized as fricative [r] when /aR/ is preceded by a consonant in front (34.3%, $\beta = -1.19, z = -2.37, p < .01$), than when preceded by a consonant in medial (20.0%) or in back (15.5%) position. For the analysis of the effect of following context, we fitted a model on those tokens, where /aR/ was not word-final (N = 360). Our data shows significant effects for Following Voice ($\beta = 1.93, z = 4.26, p < .001$): /R/ is more likely to be producing as [n] when followed by a voiced sound (91.0%) than when followed by a voiceless sound (65.3%).

3.2.4. Acoustic measures of the preceding vowel /a/

In order to investigate whether the acoustic properties of the preceding vowel /a/ is affected by the realization of /R/, we added the variables Duration, F1, F2 and F3 to a model with all significant predictors from the previous sections. Since we did not find significant effects for Duration nor for F1, our data suggests that duration and F1 are not affected by whether /R/ is realized as either fricative or trill. F2, however, is significantly lower for /a/ preceding a fricative realization of /R/ ($F_{2,mean} = 1329$ Hz) than preceding a trill (F2$_{mean} = 1345$ Hz). Finally also F3 appeared to be significantly affected by the manner of articulation of /R/: F3 is significantly higher before [n] ($F_{3,mean} = 2478$ Hz) than before [r] ($F_{3,mean} = 2434$ Hz).

Our results concerning the influence of the realization of /R/ on the preceding vowel are in line with previous findings. First, also [10] showed that the formant values of the nuclei of accented syllables depend on the place of articulation of the surrounding consonants. Second, [4] mentioned that the effect of /R/ on preceding vowels is either retraction or centralization and that palatal, dental and alveolar consonants have a high influence on F1 and F2. No previous study reports effects of the realization of /R/ on F3 of the preceding vowel (front constriction > lower F3 vs. back constriction > higher F3).

4. Concluding remarks

The aim of the present paper was two-fold: The first aim was to investigate which linguistic properties of the words predict the different realizations of /R/ in the phone sequence /aR/ in coda position in standard Austrian German. Based on 509 word tokens read by seven trained male Austrian radio speakers, we found that /R/ is least reduced in stressed syllables and in words read in isolation. Furthermore, the segmental context resulted to affect the realization of /R/: /R/ is significantly less likely to be absent when /aR/ is preceded by nasals and fricatives than when it is preceded by plosives. Also, /R/ is significantly least probable to be absent before plosives and it is most likely to be produced as a trill when followed by a vowel or by a voiced consonant. Besides these effects of prosody and segmental context, we found significant effects of morphology: /R/ tends to be more reduced (i.e., realized as [r] or deleted) when it is part of a grammatical morpheme than when it is part of the stem of a word. We discussed our finding in the light of previous findings on the effect of morphology in Dutch [22, 19].

The second aim of this paper was to investigate whether the different realizations of /R/ affect the acoustic properties of the preceding vowel /a/. The acoustic measurements of the vowel /a/ revealed that F1 and F3 are significantly affected by the realization of /R/ as either trill or fricative or when it is acoustically absent. When /R/ is absent, F1 is lower but F2 and F3 are higher than when /R/ is present. /a/ before [r] and [n] has the same average F1. Whereas F2 is lower before [r] than before [n], F3 is higher before [r] than before [n]. Moreover, /a/ tends to be longer when /R/ is absent than when it is present. These results indicate, that /R/ is not completely deleted, as cues of the /R/ are still present in the preceding vowel /a/. In the future, we need to perform perception experiments in order to test whether these observed acoustic differences are actually used by listeners.

In sum, we have shown that several factors affect the different realizations of /R/ in standard Austrian German. Our findings will inform the further development of models of pronunciation variation for human and automatic speech recognition. This study, however, is only a first step. We need more studies of the kind presented here (on different consonants and vowels, on speech from different speech styles) in order to increase our understanding of the conditions which affect pronunciation variation in standard Austrian German.

5. Acknowledgements

The work by Barbara Schuppler was supported by a Hertha-Firnberg grant from the FWF (Austrian Science Fund).
6. References


